2278

RELATION OF PREFERENCE PANEL AND TRAINED PANEL SCORES ON STERILE MILK CONCENTRATES[1, 2]

LOIS A. SATHER
Department of Food Science and Technology
AND
L. D. CALVIN
Department of Statistics, Oregon State University, Corvallis, Oregon
AND
NAOMI E. LIMING
Eastern Utilization Research and Development Division, USDA, Washington, D. C.

ABSTRACT

Two series of flavor panel tests were conducted with sterile concentrated milk samples comparing trained and preference panel scores. The samples were scored by an Eastern Utilization Research and Development Division (EURDD) Dairy Products Laboratory trained panel, an Oregon State University trained panel, and an Oregon State University (OSU) preference panel of from 140 to 170 student judges. The two trained panels used modifications of the American Dairy Science Association proposed score card for concentrated milks. The student preference panel scored the samples on the nine-point hedonic scale.

The first series consisted of 23 samples of 3:1 concentrates all from the same manufacturer and did not contain a wide range of specific flavor defects. None of the trained panels scores was highly correlated with preference scores.

The second series contained 29 samples and included 1:1, 2:1, and 3:1 sterile milks from six different manufacturers. A greater range of types and intensities of flavor defects was evident in these samples. The Dairy Products Laboratory (DPL) and OSU trained panel scores for astringent, scorched, stale, and total scores were significantly correlated with preference scores. For both trained panels, scorched and total scores gave the highest correlations with preference scores and could be used for predictive purposes.

Investigations are being conducted in a number of laboratories (1-7, 10-13) to improve the flavor and storage stability of sterile milk concentrates. The objective of these studies is to produce a sterile concentrate which will receive consumer acceptance as a beverage milk. Although trained panels are used to evaluate differences between the individual milk flavors, they do not necessarily provide information on consumer reaction to the flavors which develop on processing and storage.

A previous study in this laboratory (9) had indicated that for dry whole milk the trained panel total score or over-all evaluation was the best factor for predicting consumer preferences. This investigation was undertaken to determine if the high correlation between the trained panels total scores and the preference scores, as shown by the dry whole milk studies, might also be applicable when evaluating sterile concentrates. If the correlation is high, trained panels could be used with confidence in predicting consumer preferences when evaluating process modifications in the development of improved sterile milk concentrates.

Received for publication January 22, 1965.

[1] Technical Paper No. 1846, Oregon Agricultural Experiment Station.

[2] Presented at the 59th Annual Meeting of the American Dairy Science Association, University of Arizona, Tucson, June, 1964.

EXPERIMENTAL PROCEDURES

Two series of flavor panel tests were conducted using a trained panel at the Dairy Products Laboratory (DPL), Eastern Regional Research and Development Division, USDA, Washington, D. C., and a trained and a student preference panel at Oregon State University (OSU).

Test I consisted of 23 samples of 3:1 sterile concentrates, all manufactured by the University of Wisconsin, using varying processing procedures. Test II contained 29 samples of 1:1, 2:1, and 3:1 sterile milks supplied by six different manufacturers and stored at 40, 70, and 100 F for varying times.

On Test I, the OSU trained panel scored the

concentrates for intensity of astringent, chalky, cooked, and stale flavors on a 0, none, to 3, pronounced scale and gave each sample a total score from 1, low, to 10, high. The DPL trained panel used the American Dairy Science Association (A.D.S.A.) score-card for concentrated milk.

On Test II, both the OSU and DPL trained panels used the A.D.S.A. Proposed Score Card for Concentrated Milk for Beverage Purposes. The OSU panel used the card on the basis of a range of 1 to 10, whereas the DPL panel used a range of 31 to 40. Values of 10 and 40, respectively, were used in the absence of a flavor criticism.

The DPL panel scored all 29 concentrates in Series II twice: once when concentrates only were evaluated and the second time when samples of other types of milk also were included in the same test session.

The OSU preference panel scored all samples of both Series I and II on the nine-point hedonic scale (8), on which I indicated dislike extremely, 5, neither like nor dislike, and 9, like extremely.

For serving to the OSU panels, the concentrates were opened, stirred, then the 3:1 and 2:1 concentrates were mixed with refrigerated (40 F) distilled water in the ratio of two cans of water to one of concentrate for the 3:1 and one can of water to one can of concentrate for the 2:1 samples. All samples were mixed 1 to 2 hr before serving, stored in a household refrigerator, then served to the preference panel at refrigerator temperature, approximately 45 F. The 1:1 samples required no mixing; therefore, were refrigerated and served directly from the can. Samples for the OSU trained panel were heated in a water bath to 70 F immediately before testing. The DPL samples were prepared the day before testing and stored overnight in a household refrigerator. The DPL preparation and serving procedures were the same as those of OSU, except bottled spring water was used instead of distilled water.

Ten to twelve judges served on both the DPL and OSU trained panels and between 140 and 170 students on the OSU preference panel. The test conditions and procedures were the same as previously reported by Sather et al. (9). At OSU, each test consisted of four samples, three sterile milks and a fresh fluid milk control. At DPL, four to eight samples were included in a given test.

Mean scores were determined for each sample in each test. The score-card used by the DPL panel for Series I contained 18 flavor criticisms, but only cooked and stale were used consistently enough to include sufficient variation to be included as independent factors in the statistical analysis. The 16 other criticisms were grouped and listed as other flavors. On this score card, the judges indicated intensities by the terms slight, definite, and pronounced. For the purpose of analysis, the scores of 1, 2, and 3, respectively, were assigned to these terms. In Series II, the DPL score-card included 17 criticisms, but only astringent, cooked, scorched, and stale were used consistently enough to be analyzed separately. However, the DPL judges did, on Series II, assign a definite score value to each criticism. For example, for a slight cooked flavor instead of just marking *S*, as had been done in Series I, the judge gave a score of 39 or 38. When the judge did not assign any score to a given criticism, it was then assumed the criticism was not evident and a score of 40 was assigned for statistical purposes. There did appear to be a tendency for the judge to assign a score to only the most dominant criticisms.

Fresh fluid market milk samples were included in all tests, but the scores were not included in the statistical analysis, since these samples varied between the two laboratories.

RESULTS

The data were statistically analyzed to determine the correlation coefficients between trained and preference panel scores. The number of samples actually scored as containing a given factor, the range of mean scores, and the correlation coefficients between the trained panels and the preference scores for Series I and II are given in Tables 1 and 2, respectively.

TABLE 1

Correlation coefficients between trained and preference panel scores Series I, sterile concentrates—23 samples

Factor and panel	Samples scored as containing factor	Range of Mean scores	Corr. coeff. with preference scores
Preference panel	23	3.42- 6.90	1.00
Total score—DPL	23	33.0 -35.4	.26
Total score—OSU	23	4.17- 5.73	.30
Cooked—DPL	23	1.95- 2.44	−.12
Cooked—OSU	23	0.75- 1.75	.16
Stale—DPL	23	0.30- 1.00	.45[a]
Stale—OSU	23	0.64- 1.33	.07
Chalky—OSU	19	0- 0.25	.03
Lactone—OSU	23	0.08- 0.50	−.44[a]
Astringent—OSU	13	0- 0.33	.09
Oxidized—OSU	13	0- 0.42	−.43
Scorched—OSU	16	0- 0.33	.29

[a] Significant at 5% level.

TABLE 2

Correlation coefficients between trained and preference panel scores [a] Series II, Sterile Concentrates—29 Samples

Factor and panel	Samples scored as containing factor	Range of Mean scores	Corr. coeff. with preference scores
Preference panel	29	2.92- 5.12	1.00
Total score—DPL[b]	29	31.7 -34.9	.75**
Total score—DPL[c]	29	31.6 -34.8	.63**
Total score—OSU	29	3.2 - 6.1	.85**
Astringent—DPL[b]	20	38.7 -40.0	.51**
Astringent—DPL[c]	16	39.0 -40.0	.53**
Astringent—OSU	11	9.4 -10.0	.45**
Cooked—DPL[b]	29	34.9 -38.8	−.07
Cooked—DPL[c]	28	35.4 -40.0	.27
Cooked—OSU	29	6.7 - 9.6	−.65**
Scorched—DPL[b]	27	35.7 -40.0	.72**
Scorched—DPL[c]	26	34.8 -40.0	.62**
Scorched—OSU	20	7.6 -10.0	.81**
Stale—DPL[b]	29	31.3 -38.1	.59**
Stale—DPL[c]	29	32.6 -38.0	.54**
Stale—OSU	29	5.2 - 8.3	.39*
Other flavors—			
DPL[b]	21	36.9 -40.0	−.01
DPL[c]	16	34.7 -40.0	−.46**
OSU	10	9.2 -10.0	−.09
Caramelized—OSU	23	4.6 -10.0	.80**
Chalky—OSU	11	9.2 -10.0	.25
Lactone—OSU	28	7.1 -10.0	−.15

* Significant at 5% level.
** Significant at 1% level.
[a] For DPL, 40; and for OSU, 10, indicates no criticism. Preference panel score range was from 9, like extremely, to 1, dislike extremely.
[b] Concentrates only tested.
[c] Concentrates tested with other milk samples.

In Series I, which included only 23 samples, just two factors, DPL stale and OSU lactone, were significantly correlated with preference scores and these correlations were relatively low.

In Series II, the scoring factors for both trained panels which gave the highest correlations with preference scores were scorched and total scores. Astringent and stale scores by both trained panels were also significantly correlated with preference, but to a lesser degree.

Whether the DPL panel tested the concentrates separately or with other samples did not significantly affect the correlations, except on other flavors. When testing with other samples, other flavors gave a significant but relatively low correlation of −0.46.

Probably because of the greater range in score values and the trained panels' increased experience, Series II, containing 29 samples of concentrates, showed significant correlations between preference and trained panel scores; whereas, no high correlation was shown with the 23 samples in Series I.

As the two trained panels were using slightly different versions of the proposed A.D.S.A. Score Card, the equivalent preference scores for total scores were computed using the following regression equations:

$$Y = \text{Preference Score}$$
$$X = \text{Total Score}$$
$$\text{DPL}^{3}\ Y = -12.71 + .498x$$
$$\text{OSU}\ Y = .83 + .623x$$

The standard errors of estimate were .43 and .34, respectively, for the DPL and OSU panels. Within the range of samples used in this study Table 3 presents the DPL and OSU total scores for equivalent preference scores.

TABLE 3

Total scores for equivalent preference scores

	Disliked moderately	Disliked slightly	Neither liked nor disliked
Preference score	3	4	5
Total scores			
DPL, conc. only	31.6	33.6	35.6
DPL, with other samples	31.2	33.6	36.0
OSU	3.5	5.1	6.7

These results are in agreement with the previous study, comparing trained and preference panel scores on dry whole milk, which also indicated that the trained panels' total score was the scoring factor giving the highest correlation with preference score.

ACKNOWLEDGMENT

The authors are grateful to the several companies and to Dr. A. M. Swanson, University of Wisconsin, for furnishing the sterile milk samples. The technical advice and assistance of Dr. B. H. Webb, Chief, Dairy Products Laboratory, Joe Boyd, Statistician, USDA, Washington, D. C. and Dr. E. A. Day, Oregon State University, are gratefully appreciated.

REFERENCES

(1) Calbert, H. E., and Jackson, H. C. 1960. The Reactions of Consumers to Sterile Concentrated Milk. J. Dairy Sci., 43: 846.
(2) Day, E. A. 1963. Flavor of Milk and Concentrated Products. pp. 96-113. Proc. Sixth Milk Concentrates Conf. University of California.
(3) Leviton, Abraham. 1964. Process for Preparing Sterilized Milk Concentrates. U.S. Pat. No. 3,119,702, January 28.
(4) Leviton, A., Anderson, H. A., Vettal, H. E., and Vestal, J. H. 1963. Retarda-

[3] When concentrates only were tested.

tion of Gelation in High-Temperature Short-Time Sterile Milk Concentrates with Polyphosphates. J. Dairy Sci., 46: 310.

(5) LEVITON, A., AND PALLANSCH, M. J. 1962. High-Temperature Short-Time Sterilized Evaporated Milk. IV. The Retardation of Gelation with Condensed Phosphates, Manganous Ions, Polyhydric Compounds, and Phosphatides. J. Dairy Sci., 45: 1045.

(6) LEVITON, A., PALLANSCH, M. J., AND WEBB, B. H. 1962. Effect of Phosphate Salts on the Thickening and Gelation of Some Concentrated Milks. XVIth Intern. Dairy Congr.

(7) PATEL, T. D., CALBERT, H. E., MORGAN, D. G., AND STRONG, F. M. 1962. Changes in the Volatile Flavor Components of Sterilized Concentrated Milk During Storage. J. Dairy Sci., 45: 601.

(8) PERYAM, D. R., AND PILGRIM, F. J. 1957. Hedonic Scale Method for Measuring Food Preferences. Food Technol., (Insert 9) 11: 9.

(9) SATHER, LOIS A., TAMSMA, A., AND CALVIN, L. D. 1963. Relation of Preference Panel and Trained Panel Scores on Dry Whole Milk. J. Dairy Sci., 46: 1054.

(10) SWANSON, A. M. 1961. Processing Sterilized Concentrated Milk. Technical Product of Milk Concentrates. Univ. of Ill. Agr. Econ. Bull. 1: 9.

(11) SWANSON, A. M., AND SEEHAFER, M. E. 1963. Some Factors Involved in Sedimentation and Gelation of Concentrated Milks. pp. 56-70. Proc. Sixth Milk Concentrates Conf. Univ. of Calif.

(12) WILSON, H. K., AND HERREID, E. O. 1960. Sedimentation of Reconstituted Concentrated Milks. J. Dairy Sci., 43: 1751.

(13) WILSON, H. K., VETTER, J. L., SASAGO, K., AND HERREID, E. O. 1963. Effects of Phosphates Added to Concentrated Milks Before Sterilization at Ultra-High Temperatures. J. Dairy Sci., 46: 1038.